

(19) BUNDESREPUBLIK

DEUTSCHLAND



DEUTSCHES
PATENTAMT

(12) **Patentschrift**
(11) **DE 3519915 C2**

(21) Aktenzeichen: P 35 19 915.6-31
(22) Anmeldetag: 4. 6. 85
(43) Offenlegungstag: 4. 12. 86
(45) Veröffentlichungstag der Patenterteilung: 15. 2. 90

(51) Int. Cl. 5:

H 04 M 1/00
~~H 04 M 1/26~~
~~H 04 M 3/42~~
G 10 L 3/00
G 06 F 3/16

Innerhalb von 3 Monaten nach Veröffentlichung der Erteilung kann Einspruch erhoben werden

(73) Patentinhaber:

Telenorma Telefonbau und Normalzeit GmbH, 6000
Frankfurt, DE

(72) Erfinder:

Zinke, Joachim, Dipl.-Ing., 6000 Frankfurt, DE

(56) Für die Beurteilung der Patentfähigkeit
in Betracht gezogene Druckschriften:

DE 29 41 928 A1
CH 6 44 482

M. IMMENDÖRFER, T. HÖRMANN, D. KOPP:
»Sprachge-steuertes Telefon mit elektronischem
Telefon-register«. In: ntz Bd.37 (1984) H.8, S.496-499;
H. MULLA, J.F. VAUGHAN: »Spracherkennung und
Sprachsynthese für Dienste von Nebenstellenan-
lagen«. In: Elektrisches Nachrichtenwesen, Bd.59,
Nr.3, 1985, S.273-280;

(54) Verfahren zur Spracherkennung an Endgeräten von Fernmelde-, insbesondere Fernsprechanlagen

Beschreibung

Die Erfindung betrifft ein Verfahren zur Spracherkennung an Endgeräten von Fernmelde-, insbesondere Fernsprechanlagen nach dem Oberbegriff des Patentanspruches 1.

Aus der DE-OS 29 41 928 ist eine Schaltungsanordnung für eine Fernsprechteilnehmerstation bekannt, die eine Spracherkennungs- und Auswerteschaltung enthält, womit gesprochene Worte erkannt werden können. Bei den gesprochenen Worten handelt es sich ausschließlich um gesprochene Ziffern, die über das Mikrofon als Wahlinformation eingegeben werden. Jede einzelne gesprochene Ziffer wird dann nach eindeutiger Erkennung nach dem auf der Anschlußleitung gültigen Wahlverfahren als parallele oder serielle Einzelinformation abgegeben. Weil es sich dabei lediglich um die Wahlinformation handelt genügt es, die Lautbilder von maximal zehn Ziffern und je nach Ausführungsart von einigen wenigen Zusatzbefehlen als Referenzmuster in Speichern abzulegen. Der Speicheraufwand steigt beträchtlich, wenn weitere Befehle mit möglicherweise längeren Wortkombinationen erkannt werden sollen. Außerdem ist ein erhöhter Speicheraufwand erforderlich, wenn personenbezogene Lautbilder eingespeichert werden, so daß die betreffenden Befehle nur dann ausgeführt werden, wenn die Eingabe von einem bestimmten Sprecher aus abgegeben wird. Um eine einwandfreie Funktion zu gewährleisten, muß die geräte-individuelle Auswerteeinrichtung so beschaffen sein, daß jeder einzelne sprachlich eingegebene Befehl eindeutig erkannt werden kann. Der dazu notwendige Prozessor muß deshalb sehr aufwendig sein, damit keine Fehlinterpretationen auftreten.

Ein Verfahren zur Übermittlung von besonderen Aufträgen in Fernsprechvermittlungsanlagen, insbesondere Fernsprech-Nebenstellenanlagen, ist aus der CH-PS 6 44 482 bekannt. Hierbei handelt es sich um eine zentral in der Vermittlungsstelle angeordnete Spracherkennungseinrichtung, die einem Teilnehmer zur Verfügung gestellt wird, wenn eine besondere Steuerinformation abgegeben wird. Danach besteht die Möglichkeit, vorbestimmte Befehlsworte einzusprechen, um besondere Aufträge an das Vermittlungssystem zu geben, damit eines oder mehrere gewünschte Merkmale wirksam werden. Bei einem vorgegebenen Grad der Übereinstimmung des eingesprochenen Befehlswortes mit einem abgespeicherten Referenzmuster kann das erkannte Wort von einer Sprachausgabeeinrichtung als Quittung zum Teilnehmer gesendet werden. Da diese Einrichtung vielen Teilnehmern zur Verfügung stehen soll, muß entweder eine sprecherunabhängige Spracherkennung vorgesehen sein, oder es ist ein sehr umfangreiches Speichervolumen für teilnehmerindividuelle Speicher-einrichtungen erforderlich. Außerdem können nicht beliebig viele Spracherkennner vorhanden sein, so daß Engpässe entstehen wenn Befehlsworte nicht sofort richtig erkannt werden.

Auch in einem Aufsatz mit dem Titel: "Spracherkennung und Sprachsynthese für Dienste von Nebenstellenanlagen", der abgedruckt ist in der DE-Zeitschrift "Elektrisches Nachrichtenwesen", Band 59, Nr. 3, 1985, ab Seite 273, werden zentral angeordnete Spracherkennungseinrichtungen beschrieben. Es ist im einzelnen aufgezeigt, welche in einer Nebenstellenanlage üblichen Merkmale mittels Spracheingabe gesteuert werden können, und wie dabei zu verfahren ist. Derart umfangreiche Möglichkeiten erfordern einen relativ großen

Steuer- und Speicheraufwand, und je nach Teilnehmerverhalten treten größere Belegungszeiten auf, wenn Befehle wiederholt werden müssen. Dies kann zu Engpässen führen, wenn nicht genügend zentrale Spracherkennner zur Verfügung stehen. Es können Besetzfälle und Wartezeiten entstehen, so daß die zentralen Einrichtungen nicht immer sofort verfügbar sind.

Ein "sprachgesteuertes Telefon mit elektronischem Telefonregisterspeicher" ist in der DE-Zeitschrift ntz, Band 37, (1984, Heft 8, ab Seite 496) beschrieben. Es handelt sich dabei um eine Spracherkennungseinheit, die einem Fernsprech-Endgerät direkt zugeordnet ist. Damit sollen sprachlich eingegebene Befehlsworte und insbesondere Wahlziffern erkannt werden, so daß eine Vereinfachung der Bedienung erreicht wird. Als Vorteil wird dabei herausgestellt, daß für eine sprachgesteuerte Wahl individuelle Teilnehmernamen eingebbar sind. Davon werden auf sprecherabhängiger Basis benutzer-spezifische Referenzmuster erzeugt. Bei einem größeren Benutzerkreis tritt ein hoher Speicherbedarf auf, so daß die Referenzmuster-Daten in zusätzlichen Speichern ausgelagert werden müssen. Außerdem wird angeführt, daß der Wortschatz in Grenzen gehalten werden sollte, weil sonst mit einer erhöhten Verwechslungsgefahr und mit größeren Reaktionszeiten zu rechnen ist. Da eine solche Spracherkennungseinrichtung nur für ein Endgerät vorgesehen ist, ergeben alle Mehraufwendungen, die für einen größeren Wortschatz in Form von Speicherkapazität und für eine hohe Erkennungsgenauigkeit in Form von Rechenleistung aufzubringen sind, direkt möglicherweise zu hohe Kosten für ein derartiges Gerät.

Die Aufgabe der Erfindung besteht darin, ein Verfahren zur Spracherkennung an Endgeräten von Fernmelde-, insbesondere Fernsprechanlagen anzugeben, wobei der im Endgerät selbst notwendige Speicher- und Steueraufwand auch dann in Grenzen gehalten wird, wenn eine Vielzahl von Befehlen und Variationen der Eingabe möglich sein soll.

Diese Aufgabe wird mit Merkmalen gelöst, wie sie im Patentanspruch 1 angegeben sind.

Damit wird in vorteilhafter Weise erreicht, daß die Sicherheit bei der Spracherkennung auch dann nicht vermindert wird, wenn die Anzahl und Variation von Sprachdaten vergrößert wird, ohne daß dabei die im Endgerät vorzusehenden Speicher- und Steuereinrichtungen allzu aufwendig werden müssen. Weiterbildungen der Erfindung ergeben sich aus den Unteransprüchen.

Ein Ausführungsbeispiel der Erfindung wird nachfolgend anhand einer Zeichnung näher erläutert. Von den in einem Endgerät EG vorhandenen Einrichtungen sind nur diejenigen in Form eines Blockschaltbildes dargestellt, die mit der Erfindung unmittelbar im Zusammenhang stehen. Die von einem Mikrofon M aufgenommenen Befehle werden parallel einer Spracherkennungseinrichtung SPE und einem Sprachzwischenspeicher SZSP zugeführt. Da es sich um digitale Speichermedien handelt, werden die analogen Sprachsignale in nicht dargestellter Weise analog/digital-gewandelt. Dabei werden bekannte Codierungsverfahren zugrunde gelegt. Die Spracherkennungseinrichtung SPE vergleicht nun die angebotenen Sprachsignale mit innerhalb der Spracherkennungseinrichtung SPE abgelegten Referenzmustern, so daß festgestellt werden kann, welcher Schaltbefehl vom Benutzer des Gerätes eingegeben wurde. Die in der Spracherkennungseinrichtung SPE vorhandenen Vergleicher weisen verschiedene Schwel-

len auf, womit eine Aussage darüber abgegeben wird, ob eine Erkennung eindeutig, unsicher oder fehlerhaft ist.

Wird die Bedeutung eines Wortes oder einer Wortfolge eindeutig erkannt, weil sie mit einem vorgegebenen Referenzmuster ziemlich vollkommen übereinstimmt, so erscheint ein Signal *S 1*, das die sofortige Ausführung des betreffenden Befehls zur Folge hat und außerdem einer Anzeigeeinrichtung *AE* zugeführt wird. Selbstverständlich ist es auch denkbar, die Ausführung des Befehles von einem weiteren Befehl abhängig zu machen, wenn der Benutzer erkannt hat, daß die Spracherkennungseinrichtung *SPE* die gewünschte Funktion richtig erkannt hat, falls für besondere Betriebsfälle eine erhöhte Sicherheit verlangt wird.

Falls die Vergleichseinrichtungen der Spracherkennungseinrichtung *SPE* keine eindeutige Übereinstimmung zwischen einem eingegebenen Sprachsignal und den abgelegten Referenzmustern feststellen können, so wird ein besonderes Signal *S 2* abgegeben. Dieses Signal gelangt in eine Steuereinrichtung *SE*, wodurch die Anschlußleitung *AL* belegt wird. Es wird ein Schaltmittel *K 2* geschlossen, wodurch der Sprachzwischenspeicher *SZSP* mit der Anschlußleitung *AL* verbunden wird. Von der Spracherkennungseinrichtung *SPE* aus wird veranlaßt, daß die im Sprachzwischenspeicher *SZSP* eingespeicherte Information ausgelesen wird. Die im Sprachzwischenspeicher *SZSP* abgelegte Sprachinformation gelangt je nach Betriebsart der Anschlußleitung *AL* entweder direkt digital zur Fernsprechanlage *FA*, oder sie wird bei analoger Anschlußleitung *AL* zuvor analog gewandelt. Bei digitaler Anschlußleitung *AL* wird diese Information in einem sogenannten *B*-Kanal als Nutzinformation übertragen. Wenn ein Endgerät an das digitale Fernmeldenetz zur integrierten Übertragung mehrerer Dienste (*ISDN*) angeschlossen ist und dabei zwei sogenannte *B*-Kanäle zur Verfügung stehen, kann die Übertragungszeit auf die Hälfte reduziert werden, indem beide *B*-Kanäle benutzt werden.

In der Fernsprechanlage *FA* befindet sich eine zentral angeordnete Spracherkennungseinrichtung mit erhöhter Speicher- und Rechenkapazität. Die über die Anschlußleitung *AL* eingegebene Sprachinformation kann dann mit wesentlich mehr Referenzmustern verglichen werden, so daß mit einer erhöhten Wahrscheinlichkeit der gewünschte Befehl eindeutig identifiziert werden kann. Die von der zentralen Spracherkennungseinrichtung daraufhin gewonnene Erkenntnis wird über die Anschlußleitung *AL* zurück zum Endgerät *EG* übertragen. Bei digitaler Anschlußleitung *AL* bietet sich hierfür der Signalisierungskanal *D* an. Die von der Fernsprechanlage *FA* kommende Information gelangt in die Steuereinrichtung *SE* des Endgerätes *EG*, wobei dann die Anzeigeeinrichtung *AE* angesteuert wird. Handelt es sich um einen Befehl, der innerhalb der Fernsprechanlage *FA* auszuführen ist, so kann dies unmittelbar geschehen, indem die dortige zentrale Steuerung angereizt wird. Bei innerhalb des Endgerätes *EG* auszuführenden Befehlen, wird die Steuereinrichtung *SE* des Endgerätes *EG* tätig.

Natürlich ist es auch denkbar, daß selbst die zentrale Spracherkennungseinrichtung einen Befehl nicht eindeutig erkennen kann. Es wird dann auf die zuvor beschriebene Weise eine Fehlermeldung über die Anschlußleitung *AL* zum Endgerät *EG* übertragen. Die Steuereinrichtung *SE* nimmt diesen Befehl auf und steuert die Anzeigeeinrichtung *AE* in entsprechender Weise an. Der Benutzer erkennt dann an der Anzeige, daß die

von ihm eingegebene Sprache nicht richtig erkannt worden ist.

Es ist außerdem vorgesehen, daß bei Spracheingaben, die sehr weit abweichen von den vorgesehenen Referenzmustern von der Spracherkennungseinrichtung *SPE* des Endgerätes *EG*, direkt ein Fehlersignal *F* abgegeben wird. Dieses Signal gelangt direkt zur Anzeigeeinrichtung *AE*, ohne daß eine Funktion bewirkt wird, und ohne daß Sprachinformationen über die Anschlußleitung *AL* zur Fernsprechanlage und zur dortigen zentralen Spracherkennungseinrichtung übertragen werden.

Um den Datenaustausch zwischen dem Endgerät *EG* und der zentralen Spracherkennungseinrichtung in der Fernsprechanlage *FA* zu vereinfachen, und die dafür benötigte Zeit zu verkürzen, kann im Endgerät *EG* eine Sprachvorverarbeitungseinrichtung *SPV* vorgesehen sein. Dann werden die im Sprachzwischenspeicher *SZSP* befindlichen Sprachinformationen von der Sprachvorverarbeitungseinrichtung *SPV* so aufbereitet, daß die über die Anschlußleitung *AL* zu übertragende Information direkt geeignet ist, die in der Fernsprechanlage *FA* angeordnete Spracherkennungseinrichtung, bzw. deren Speicher gezielt anzusteuern. Wenn das Endgerät *EG* mit einer Sprachvorverarbeitungseinrichtung *SPV* ausgestattet ist, so wird von der Steuereinrichtung *SE* anstelle des Schaltmittels *K 2* das Schaltmittel *K 1* betätigt, so daß der Ausgang der Sprachvorverarbeitungseinrichtung *SPV* mit der Anschlußleitung *AL* verbunden ist.

Wenn der Benutzer an der Anzeigeeinrichtung *AE* erkennt, daß eine Fehlermeldung *F* erscheint, so kann durch Eingabe in deutlicherer Form der Befehl wiederholt werden. Bei der Anzeige ist außerdem denkbar, daß zusätzlich eine Information erscheint, wenn die in der Fernsprechanlage *FA* befindliche zentrale Spracherkennungseinrichtung in Anspruch genommen wurde. Der Benutzer wird dann veranlaßt, sich eine präzise Abgabe der Sprachinformation anzugewöhnen.

Patentansprüche

1. Verfahren zur Spracherkennung an Endgeräten von Fernmelde-, insbesondere Fernsprechanlagen, wobei innerhalb des Endgerätes eine Spracherkennungseinrichtung (Spracherkennung) angeordnet ist, die Sprachsignale vom Mikrofon aufnimmt und daraus Befehle zur Steuerung von Merkmalen oder für den Aufbau von Verbindungen bildet, indem ein Vergleich durchgeführt wird zwischen als Referenzmuster abgespeicherten Befehlsworten und dem von der jeweiligen Spracheingabe erzeugten Referenzmuster, und wobei das Endgerät an eine Vermittlungsanlage angeschlossen ist, die eine zentral angeordnete Spracherkennungseinrichtung aufweist, dadurch gekennzeichnet,
daß im Endgerät (*EG*) ein besonderer Sprachzwischenspeicher (*SZSP*) vorgesehen ist, der die dem Spracherkennung (*SPE*) angebotenen Sprachsignale zusätzlich aufnimmt,
daß die zentrale, in der Fernmelde- oder Fernsprechanlage (*FA*) angeordnete Spracherkennungseinrichtung mit erhöhter Speicher- und Rechenkapazität automatisch angesteuert wird, und die im Sprachzwischenspeicher (*SZSP*) befindliche Information dorthin übertragen wird, wenn im Endgerät (*EG*) befindliche Spracherkennungseinrichtung (*SPE*) eine Spracheingabe nicht eindeutig

- erkennen und einem der abgespeicherten Referenzmuster zuordnen kann.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die zur zentralen Spracherkennungseinrichtung zu sendenden Sprachsignale in analoger Form übertragen werden. 5
3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die zur zentralen Spracherkennungseinrichtung zu sendenden Sprachsignale in digitaler Form (z. B. PCM) übertragen werden. 10
4. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß dem Sprachzwischenspeicher (SZSP) eine Sprachvorverarbeitungseinrichtung (SPV) nachgeschaltet ist, und der zentralen Spracherkennungseinrichtung anstelle von direkten Sprachsignalen analysierte Ansteuerdaten zugeführt werden. 15
5. Verfahren nach Anspruch 4, dadurch gekennzeichnet, daß die Analyse der Spracheingabe durch die Sprachvorverarbeitung (SPV) zur gleichen Zeit 20 vorgenommen wird, zu der die Spracherkennungseinrichtung (SPE) arbeitet.
6. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die von der im Endgerät (EG) befindlichen Spracherkennungseinrichtung (SPE) oder 25 von der zentralen Spracherkennungseinrichtung erarbeitete Information zur Kontrolle einer Anzeigeeinrichtung (AE) zugeführt wird.
7. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß beim Vergleicher in der im Endgerät 30 (EG) befindlichen Spracherkennungseinrichtung (SPE) zwei Ansprechschwellen vorgesehen sind, wobei die erste Schwelle (S_1/S_2) unterscheidet zwischen einer sicheren und einer unsicheren Erkennung von Spracheingaben und eine zweite 35 Schwelle (S_2/F) zwischen einer unsicheren Erkennung und einer fehlerhaften Eingabe entscheidet.
8. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß ein Unterschreiten der zweiten Schwelle (S_2) als fehlerhafte Spracheingabe interpretiert wird, wobei ein Fehlersignal (F) erzeugt 40 wird, welches ohne eine Funktion zu bewirken lediglich in der Anzeigeeinrichtung erscheint, wobei dann eine Ansteuerung der zentral angeordneten Spracherkennungseinrichtung unterbleibt. 45

Hierzu 1 Seite(n) Zeichnungen

- Leerseite -

